Conducting neuropsychological tests with a humanoid robot: design and evaluation

Duc-Canh Nguyen Gérard Bailly Frédéric Elisei

GIPSA-lab, CNRS & Grenoble-Alpes Univ.
Saint Martin d'Hères, France
{first_name.family_name}@gipsa-lab.fr

***Abstract*— Socially assistive robot with interactive behavioral capability have been improving quality of life for a wide range of users by taking care of elderlies, training individuals with cognitive disabilities or physical rehabilitation, etc. While the interactive behavioral policies of most systems are scripted, we discuss here key features of a new methodology that enables professional caregivers to teach a socially assistive robot (SAR) how to perform the assistive tasks while giving proper instructions, demonstrations and feedbacks. We describe here how socio-communicative gesture controllers – which actually control the speech, the facial displays and hand gestures of our iCub robot – are driven by multimodal events captured on a professional human demonstrator performing a neuropsychological interview. Furthermore, we propose an original online evaluation method for rating the multimodal interactive behaviors of the SAR and show how such a method can help designers to identify the faulty events.**

Keywords-- socially assistive robot; humanoid robot; multimodal behavior; subjective evaluation;

I. INTRODUCTION

A. A socially assistive rotot

A socially assistive robot is an assistive robot aiding people through social interactive behaviors rather than physical interaction [25]. Several socially assistive robot (SAR) systems have been proposed and designed to engage people into various interactive exercises such as physical training [1], neuropsychological rehabilitation [2] or cognitive assistance [3]. Depending on the objectives of the Human-Robot Interaction (HRI), SAR faces different challenges. One of the important dimensions is the length of this interaction. Several SAR are concerned with long-term interaction, aiming at providing a single user with social glue and affective relations. Paro [4] is emblematic of that challenge (see Leite et al [5] for a review). These companion robots often play the role of pets or majordomos. In contrast, several SAR have also been designed to engage into short-term task-oriented interactions. The challenge is here more oriented towards attention and quick adaptation.

Our works focus on the development of socio-communicative abilities of a SAR for short-term interactions. Particularly, in this paper, we present the SOMBRERO framework which aims at providing a humanoid robot with multimodal interactive behaviors – such as speech, gaze arm gestures, etc. – in order to perform a neuropsychological test, demonstrated by professionals.

B. SOMBRERO framework

The three main steps of learning interaction by demonstration are given in Figure 1: we should (1) collect representative interactive behaviors from human coaches; (2) build comprehensive models of these overt behaviors and a priori knowledge (task & user model, etc.); and then (3) provide the target robot with appropriate gesture controllers to execute the desired behaviors.

Most interaction models embedded into HRI systems are strongly inspired by Human-Human Interactions (HHI), if not entirely trained on HHI data. This HHI-based framework faces several problems: (1) the scaling of the human model to the interaction capabilities of the robots in terms of physical limitations (degrees of freedom) and perception, action and reasoning; (2) the drastic changes of human behaviors in front of robots or virtual agents [6]; (3) the modeling of joint interactive behaviors (4) the replay and assessment of these behaviors by the robot.

SOMBRERO proposes to solve the two first issues by enabling coaches to demonstrate human-robot interaction (HRI) via immersive teleoperation, i.e. by direct robotic

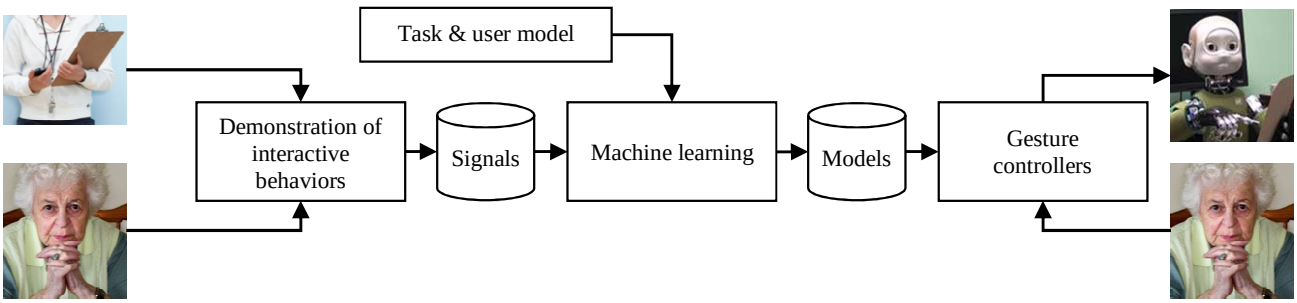

Figure 1: The main steps of learning interaction by demonstration

embodiment. HRI is thus bootstrapped by a robot-mediated HHI, which may be broadly considered as a cognitive infocommunication system (see section IV).

The immersive teleoperation of the gaze and lip movements of our iCub robot Nina is described in Guillermo et al [7]. This technique artificially provides the SAR with cognitive – notably social – skills that are intrinsically adapted to the robot's sensorimotor abilities. Moreover subjects may have the impression to interact with a truly autonomous SAR and thus provide the interaction model with genuine behavioral data.

The third issue – i.e. learning–based modelling of interactive behaviors – is an emerging concept (see pioneer work performed by Otsuka et al [8] or more recently [9]). While learning by demonstration [10] has been quite effective for sensorimotor tasks such as walking or grasping where the robot interacts with the physical environment, its extension to HRI presupposes HRI data that are difficult to acquire. We proposed [11] [12] to train statistical behavioral models that encapsulate discrete multimodal events performed by the interlocutors into a unique dynamical system that could be further used to monitor behaviors of one interlocutor and generate behaviors of the other.

We address here the fourth issue i.e. the replay and assessment of interactive behaviors by the robot.

II. THE CURRENT CONTRIBUTION

We should in fact verify that the multimodal behaviors planned by the interaction model can effectively be reproduced by the target robot and that these multimodal behaviors are perceived as adequate by human interlocutors.

A. *The scenario*

Our interviews are based on the French adaptation [13] of the Selective Reminding Test [14] named the RL/RI 16. It provides a simple and clinically useful verbal memory test for identifying loss of episodic memory in the elderly.

The RL/RI 16 protocol consists in four phases. The first one is the progressive learning of 16 words together with their semantic categories four by four. This phase includes two main tasks: (1) item identification (the interviewer displays four items and asks the subject spell out each name by giving its category); (2) immediate recall of items (the subject should recall each item by its category while the items are hidden). The second phase is three successive recall tasks (i.e. free recall, complemented by an indexed-by-category recall for the unrecovered items) separated with a distractive task (reverse counting). In the next, there is a recognition task involving the 16 items, 32 distractors (16 different words with the same semantic category and 16 true distractors) and (4) a delayed free and indexed recall (not administrated in the present study). Mnesic performance is evaluated by comparing recall rates of the subject with regards to mean & standard deviations observed within sane control population of the same age interval.

Most professionals use folders with sheet of papers. In order to avoid complex dexterous gestures to be performed by the SAR, we adopted a modified scenario using two tablets: one tablet facing the robot to score the subject's answers and the other tablet facing the subject to display visual stimuli i.e. items to be learned or recognized.

B. *Interactive data*

The demonstrations used here have been performed by a female professional neuropsychologist. Since the immersive teleoperation of the upper body (notably of the arms) is not available yet, the discrete multimodal events have been collected via semi-automatic labeling of HHI. The motion of 25 reflective markers placed on the plexus, shoulders, head, arms, indexes and thumbs of the professional interviewer were monitored thanks to a Qualysis® system with 4 cameras. A Pertech® head-mounted monocular eye tracker also monitors the gaze of the interviewer (see Figure 2). Speech data are captured via OKMII high-quality ear microphones and are recorded synchronously with a side-view video by HD camera.

Each interview lasts around 20", comprising the collection of personal records, the core RL/RI protocol and final report of performance. We analyze here a total two hours of multimodal data for five subjects, interacting with the professional interviewer we used as unique demonstrator of the appropriate socio-communicative behaviors for that particular task.

Figure 2. Visual data. Left: head-related view from the eyetracker scene camera. The marker superimposed to the scene camera features the current gaze fixation point. Right: side view from a fixed HD camera.

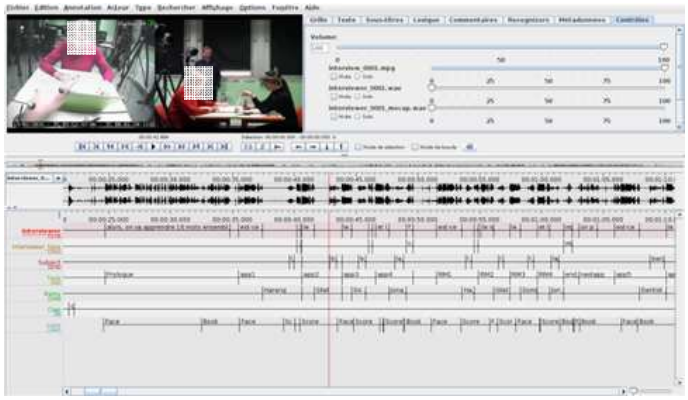

Figure 3. Labeling gaze & speech events with Elan.

C. *Gestural scores*

Elan [15] (see Figure 3) and Praat [16] were used to semi-automatically identify speech, gaze and arm events. The behavioral model has to orchestrate these events according to the different sub-tasks and should be able to generate motor actions from percepts. Modality-specific gesture controllers have then to reproduce final motions from these discrete motor events. Our hypothesis is that the compositional richness of

such discrete events is able to capture the diversity of multimodal behaviors.

D. Gesture controllers

Speech. We transcribed speech and aligned its phonetic content with the acoustic signals uttered by both the interviewer and the subjects. We mainly spot items in the subject's speech in order to trigger scoring and interviewer's feedbacks. The interviewer's speech was analyzed more in-depth with a special attention to prosody and in particular to backchannels [17]. The orthographic transcription of her discourse augmented with prosodic markers and breath noises is then played by the audiovisual text-to-speech synthesizer controlling Nina's loudspeaker and facial movements [18].

Arm gestures. While the human interviewer was displaying word items and scoring using sheets of paper, we decided to use tablets to display items and pretend to trigger the display and take notes (see Figure 4). In fact, subjects project human skills and capabilities onto agents – including mnesic capabilities – and would be very disturbed if the artificial interviewer does not take any notes despite the fact that its processing power does not require such a physical display. Such a behavior is thus clearly imposed by social rules.

Arm displacements and finger clicks are programmed to trigger display on the subject's tablet (show/hide items) and take notes (monitor correct responses). The arm gesture controller uses the iCub Cartesian Interface [19], which enables the control of the robot's arm directly on operational space by providing the desired position and orientation of one end-effector (here the index finger of the right hand). The arm controller also provides task-specific movements: preparing to click, clicking, and going back to rest position. Figure 4 illustrates the position of robot's right arm while scoring and resting. In the experiment, the left arm remains fixed. It holds the scoring tablet, while the right arm movements are adapted so as to follow the timing of the writing gestures of the human interviewer.

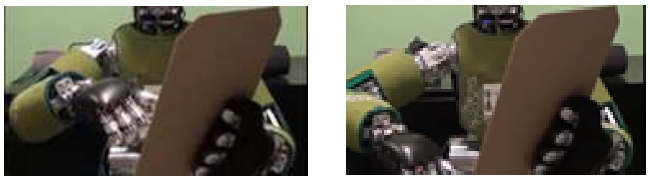

Figure 4. Robot's arm while (left) scoring and (right) resting.

Gaze. We distinguish three main regions of interest of the interviewer's gaze: (1) the subject's face; (2) the scoring tablet (i.e. the scoring sheet and chronometer used in the original HHI); (3) the subject's tablet (i.e. the notebook used in HHI demonstrations). Note that all arm gestures are performed with visuomotor supervision: since robot motion is often slower than human motion, all arm motions are preceded by one fixation towards the target if any and accompanied by gaze smooth pursuit till completion. This visuomotor supervision supersedes the original fixation patterns.

The gaze gesture controller uses the iCub gaze controller [20], which provides direct control of saccades, fixations and smooth pursuit while implementing the binoccular vergence, the oculo-collic and vestibule-ocular reflexes. These gestures can be performed by a parametrized combination of neck and eyes movements. For simplicity, the Cartesian gaze controller is provided with the 3D position of the current region of interest and a fixed contribution of neck movements of 50%. Figure 5 presents robot in two positions: (a) look at subject's face, (b) look at scoring tablet.

Eyelids. Although we did not track eyelids' movements, we developed a specific eyelids gesture controller in order to provide Nina's behavior with additional socio-communicative cues such as blinking as well as redundant cues such as the coupling of eyelids aperture with eyes elevation [21] [22] and speech articulation [23]. Figure 6 illustrates the coupling of eyelids aperture with eyes elevation.

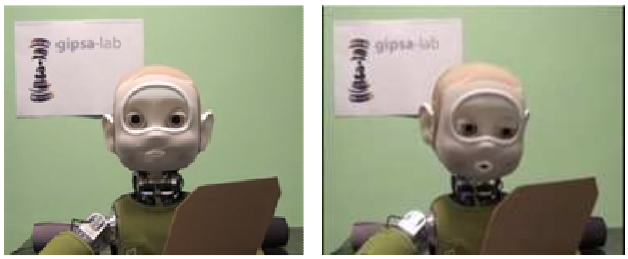

Figure 5. Robot's gaze looking at subject's face and scoring tablet.

Figure 6. Robot's eyelids with gaze looking down (left) or straight (right).

III. Evaluation

These complex and coordinated behaviors should be perceived and interpreted correctly by subjects. We have shown that the morphology and appearance of effectors can strongly impair the perception of planned gestures [14]. We thus ask third parties to rate the final rendering of our HHI multimodal score by our robotic embodiment in order to check if the reconstructed behavior of our human demonstrator is still relevant and if the mapping between discrete events and gestures are correctly performed by our gestural controllers.

A. State of the art

Most subjective evaluations of HRI behavior have been performed using questionnaires, where subjects or third parties are asked to score specific dimensions of the interaction on Likert scale. For example, Fasola et al [1] rated several aspects such as pleasure, interest, satisfaction, entertainment and excitation of a SAR monitoring physical exercises. Huang et al [9] assessed a narration humanoid robot along several dimensions such as immediacy, naturalness, effectiveness, likability and credibility. Zheng et al [24] compared control strategies for robot arm gestures along dimensions such as intelligibility, likeability, anthropomorphism and safety.

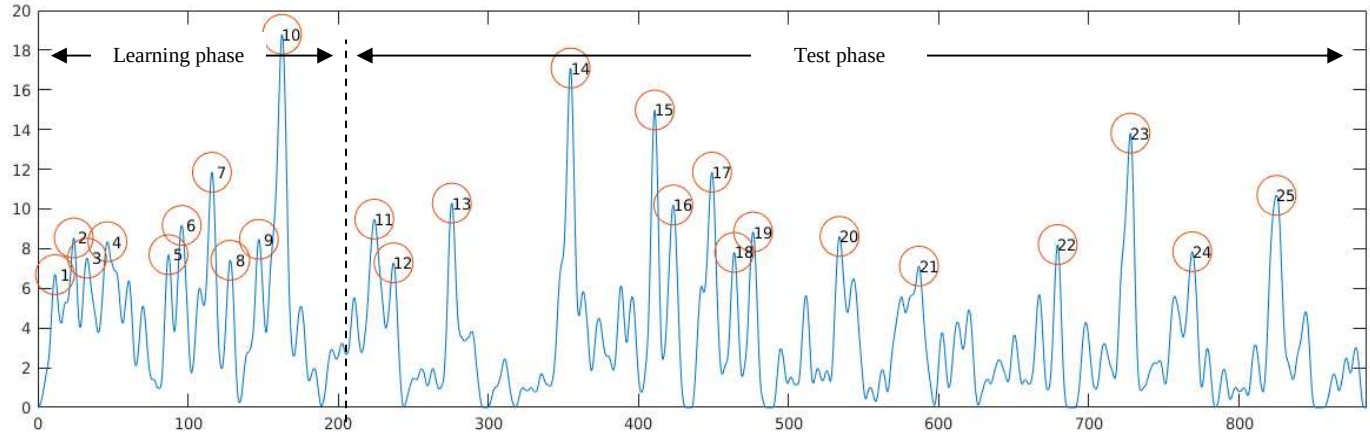

Figure 7. Density of yuck responses for our replayed interaction. Each yuck response is weighted by a Hanning window of 5s in order to smooth the density of responses and overlapped/added to the others. Maxima of this time-dependent histogram reveal multimodal behaviours that are judged inadequate by a majority of raters. The 25 main maxima are commented in the text.

Although delivering very useful information notably for sorting between competing control policies or settings, these questionnaire-based evaluations provide developers with poor information about faulty behaviors: the evaluation is in fact performed off-line and questions mostly address global properties of the entire interaction.

B. *Designing and performing an on-line evaluation*

In line with online evaluation methods deployed for audio [25] and video [26], we designed an on-line evaluation technique that consists in asking raters to immediately signal faulty behaviors when they observe them. We will use the term "*yuck response*" used by de Kok [27] for evaluating the adequacy of automatically generated backchannels.

Since raters cannot both experience and rate an interaction, we ask them to put themselves in the place of subjects who have previously experienced the interaction. The technique thus consists in replaying a recorded interaction previously performed by a Guinea pig and asking subjects to rate the adequacy of the SAR's behavior with regards to the Guinea pig's verbal behavior. Note that this on-line evaluation task can also be performed by the Guinea pigs who have previously experienced the interaction.

This online evaluation can be performed in front of the robot itself – with the benefit of physical presence but the challenge of coping with the active perception of each rater who may change his/her viewing position – or in front of a video recorded from the Guinea pigs' perspective. We adopt here the latter option, i.e. we filmed the robot replaying the situated interaction using a camera placed approximatively at the mean position of the Guinea pigs' eyes.

We created a website[1] where we ask people to press the "ENTER" key anytime they feel the robot behavior is incorrect. This provides a time-varying probability density function of incorrect behaviors. The maxima of the density function provide *when cues*, i.e. time-intervals for which a majority of raters estimate the behavior is inappropriate or hinders the interaction. Further diagnostic of *what cues* cause these faulty behaviors has to be performed by roboticists and system designers. This on-line evaluation task is preceded by a quick screening of subjects (age, sex and mother tongue) and a familiarization exercise, and followed by a questionnaire that asks the subjects' judgements (five-level Likert) on nine points:

1. Did the robot adapt to the subject?
2. Did the subject adapt to the robot?
3. Did you feel relaxed?
4. Did you feel secure?
5. Was the rhythm of the robot's behavior well adapted?
6. Was the interaction pleasant?
7. Was the multimodal behavior appropriate?
8. Did the robot pay attention while speaking?
9. Did the robot pay attention while listening?

C. *Results*

We report here the objective and subjective scores of 53 French natives, who performed the entire evaluation. 29 are males and 24 females. The average age of the participants is 32 ±12 years. Figure 7 shows the weighted cumulated yuck responses of our participants. Most maxima of this density function have clear interpretations. We further used Elan (see Figure 8) to associate these major yuck responses with multimodal events. Here are the 25 most signaled events:

- 1, 4, 6 & 7: the robot here performs clicking gestures on its tablet to show or hide items onto the subject's display that was not available to raters. Such ungrounded gestures are thus perceived as distractors by subjects. These yuck responses are located at the beginning of the interview, during the learning phase.
- 2, 3, 5, 9, 17, 20, 21, 24 & 25: participants also detected that gaze towards the subject was missing or too much delayed with reference to the interviewee's answers to questions or when delivering instructions. While such a behavior is quite legible when performed by the interviewer – who did not want to interfere with the subject's thoughts – it seems completely unacceptable when performed by a SAR, whose intentions are much less readable.

[1] http://www.gipsa-lab.fr/~duccanh.nguyen/assessment/

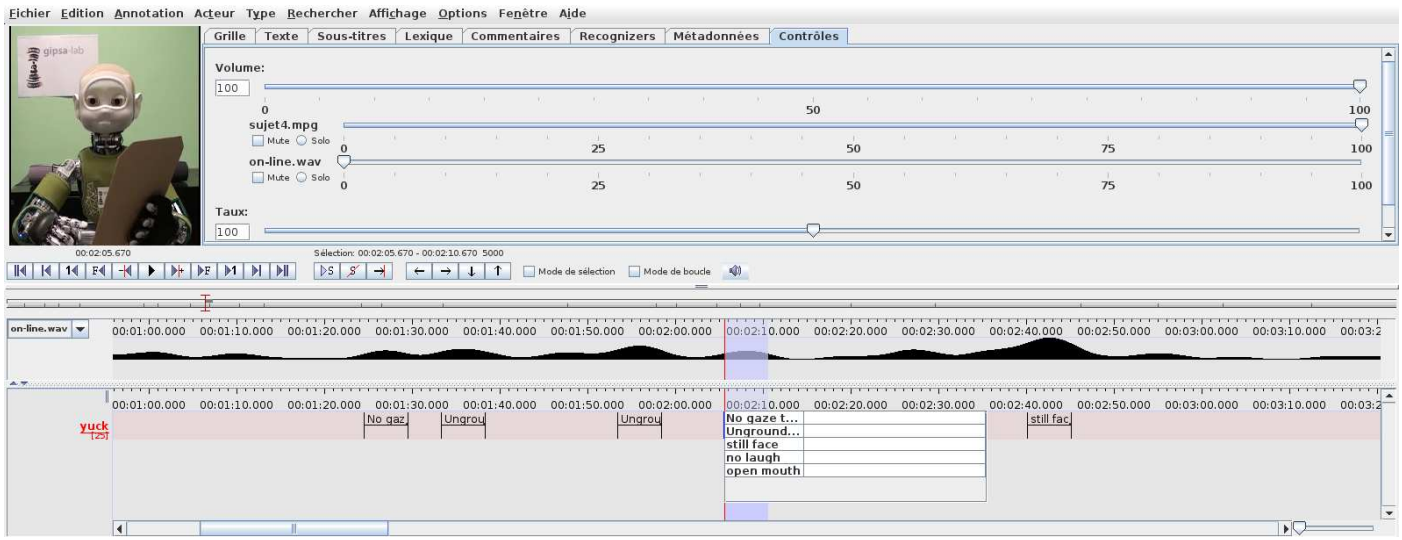

Figure 8. Using ELAN [15] to diagnose the possible causes of major yuck responses.

- 10, 11, 12, 13, 16, 18, 19 & 23: the robot remains still – with the exception of quasi-periodic blinks – for too long, notably during periods of poor interactive activity of the interviewee such as reverse counting or covert thinking. This absence of input observations results in no generated movements. We are currently working on a more lively default listening behavior that will cope with periods of poor external stimulation.
- 14 & 15: these particular misbehaviors are explained by the persistence of a large mouth opening well after finishing speaking. This failure is now identified and has been corrected: it was due to a faulty audiovisual segment that was improperly articulated during a silent pause.
- 22: In several places, the subject joked and laughed. The lack of SAR response to this strong call for social support during episodes of embarrassment is here rightly penalized by raters.
- 8: We did not found any obvious explanation for this particular yuck response in the learning phase.

Figure 9 shows the global subjective judgements performed at the end of the online evaluation. Raters strongly agree that the robot displays a rather decent behavior with regards to the task it is assigned to. This noteworthy evaluation is confirmed by the free comments given before closing each session: nearly all raters declared to be impressed by the overall quality and relevance of the robot's behaviors. Three other criterions also reach a large consensus: the robot pays attention to her interlocutors when speaking, the SAR is perceived as secure and adapts to its interlocutors. Three dimensions receive a lower consensus: attention during listening is impaired by several failures that have been identified and commented above while the interaction is judged not so relaxed and pleasant. We also need to work out the speed: several raters reported that speech intelligibility would have been improved by a lower speech rate at the beginning of the interview.

In the free comments, several raters mention the rather directive style of our female interviewer and the absence of emotional vocal and facial displays on our SAR – e.g. laughs and smiles. We plan to add such segments in our audiovisual speech repository in complement to breath noises and humming. We will also urge the design of articulated eyebrows.

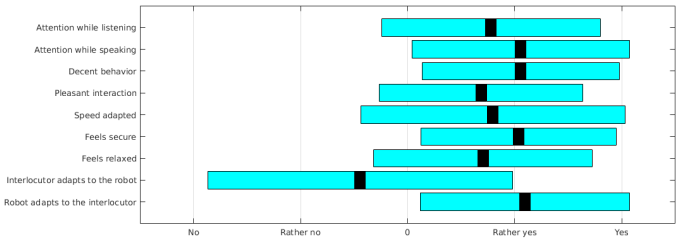

Figure 9. Global subjective judgements.

IV. COGNITIVE ROBOTICS AND INFOCOMMUBNICATIONS

Experienced professional pilots here transfer their task-specific socio-communicative skills to a SAR by mediating their usual HHI practice though a robotic embodiment. No doubt that this mediation strongly affects behaviors of pilots and subjects. We face here a challenging situation where pilots have to co-evolve both with SAR expressive and performative capabilities and with the adaptation of their human interlocutors to the affordances of the artificial agent. These themes have a strong relevance to the field of cognitive infocommunications, notably because cognitive robots are right at the crossroads of technology and humankind: pilots, subjects and SAR build an unprecedented ecosystem that might benefit from a long-term co-evolution in order to ease short-term interactions.

V. CONCLUSIONS

We here gave an overview of the SOMBRERO framework for collecting, modeling, controlling and evaluating SAR. All the building blocks are almost operational and have been

evaluated separately. We notably put forward an original framework for the online evaluation of interactive behavior that offers subsequent glass-box assessment: post-hoc reverse engineering should be performed by the SAR designers to identify the potential causes of the most consensual yuck responses. We will verify that the correction are well accepted and do not generate new errors.

One of the major missing features of our current control policy concern head movements that are now just contributing to the gaze direction of the SAR, i.e. the head remains still is no gaze shifts are programmed. In the same vein as our eyelids' controller, we plan to combine contributions of head movements to gaze with their co-verbal contributions – notably as encoders of audiovisual prosody.

In the mid-term, we plan to conduct robot-mediated HHI using immersive teleoperation very soon and see what parts of this framework should be corrected. One of the key challenges is the system's adaptation. Mihoub et al [11] have shown that a subject-independent gaze model may be parameterized to adapt to specific social profiles. We will see if this approach scales to multimodal behavior planning and control.

VI. Acknowledgements

This research is supported by the ANR (ANR-14-CE27-0014). We thank Ghatsan Hasan for taking care of Nina, Alessandra Juphard for conducting the neuropsychological tests, our five elderly subjects and the crowdsourced raters for their patience and involvement.

VII. Appendix

The multimodal data and label files are freely available at: www.gipsa-lab.fr/projet/SOMBRERO/data

References

[1] J. Fasola and M. Mataric, "A socially assistive robot exercise coach for the elderly," *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 3–32, 2013.

[2] J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied, "Why Robots? A survey on the roles and benefits of social robots in the therapy of children with autism," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 593–618, 2013.

[3] M. E. Pollack, L. Brown, D. Colbry, C. E. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, and I. Tsamardinos, "Autominder: An intelligent cognitive orthotic system for people with memory impairment," *Robotics and Autonomous Systems*, vol. 44, no. 3, pp. 273–282, 2003.

[4] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," presented at the Proceedings of the 2005 IEEE international conference on robotics and automation, 2005, pp. 2785–2790.

[5] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[6] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Autonomous Robots*, vol. 24, no. 2, pp. 159–178, 2008.

[7] G. Guillermo, P. Carole, F. Elisei, F. Noel, and G. Bailly, "Qualitative assesment of a beaming environment for collaborative professional activities," in *European conference for Virtual Reality and Augmented Reality (EuroVR)*, 2015, p. 8 pages.

[8] K. Otsuka, Y. Takemae, and J. Yamato, "A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head directions, and utterances," in *International Conference on Multimodal Interfaces (ICMI)*, Seattle, WA, 2005, pp. 191–198.

[9] C.-M. Huang and B. Mutlu, "Learning-based modeling of multimodal behaviors for humanlike robots," in *ACM/IEEE international conference on Human-Robot Interaction (HRI)*, 2014, pp. 57–64.

[10] S. Calinon, F. Dhalluin, E. Sauser, D. Caldwell, and A. Billard, "Learning and reproduction of gestures by imitation: An approach based on Hidden Markov Model and Gaussian Mixture Regression," *IEEE Robotics and Automation Magazine*, vol. 17, no. 2, pp. 44–54, 2010.

[11] A. Mihoub, G. Bailly, and C. Wolf, "Learning multimodal behavioral models for face-to-face social interaction," *Journal on Multimodal User Interfaces (JMUI)*, vol. 9, no. 3, pp. 195–210, 2015.

[12] A. Mihoub, G. Bailly, C. Wolf, and F. Elisei, "Graphical models for social behavior modeling in face-to face interaction," *Pattern Recognition Letters*, vol. 74, pp. 82–89, 2016.

[13] M. Van der Linden, F. Coyette, J. Poitrenaud, M. Kalafat, F. Calicis, C. Wyns, and S. Adam, "L'épreuve de rappel libre / rappel indicé à 16 items (RL/RI-16)," in *L'évaluation des troubles de la mémoire : présentation de quatre tests de mémoire épisodique avec leur étalonnage*, M. Van der Linden, Ed. Marseille, France: Solal, 2004, pp. 25–47.

[14] E. Grober and H. Buschke, "Genuine memory deficits in dementia," *Developmental Neuropsychology*, vol. 3, pp. 13–36, 1987.

[15] B. Hellwig and D. Uytvanck, "EUDICO Linguistic Annotator (ELAN) Version 2.0.2 manual," Max Planck Institute for Psycholinguistics, Report, 2004.

[16] P. Boersma and D. Weenink, *Praat, a System for doing Phonetics by Computer, version 3.4*. 1996.

[17] Bailly, Gérard, Elisei, Frédéric, Juphard, Alexandra, and Moreau, Olivier, "Quantitative analysis of backchannels uttered by an interviewer during neuropsychological tests," in *Interspeech*, San Francisco, CA, 2016.

[18] Parmiggiani, Alberto, Elisei, Frédéric, Maggiali, Marco, Randazzo, Marco, Bailly, Gérard, and Metta, Giorgio, "Design and validation of a talking face for the iCub," *International Journal of Humanoid Robotics*, vol. 12, no. 3, p. 20 pages, 2015.

[19] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots," presented at the Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, 2010, pp. 1668–1674.

[20] A. Roncone, U. Pattacini, G. Metta, and L. Natale, "A Cartesian 6-DoF Gaze Controller for Humanoid Robots," presented at the Robotics: Science and Systems (RSS), 2016.

[21] F. Foerster, G. Bailly, and F. Elisei, "Impact of iris size and eyelids coupling on the estimation of the gaze direction of a robotic talking head by human viewers," in *Humanoids*, Seoul, Korea, 2015.

[22] G. Bailly, F. Elisei, S. Raidt, A. Casari, and A. Picot, "Embodied conversational agents : computing and rendering realistic gaze patterns," in *Pacific Rim Conference on Multimedia Processing*, Hangzhou - China, 2006, vol. LNCS 4261, pp. 9–18.

[23] P. Badin, G. Bailly, L. Revéret, M. Baciu, C. Segebarth, and C. Savariaux, "Three-dimensional linear articulatory modeling of tongue, lips and face based on MRI and video images," *Journal of Phonetics*, vol. 30, no. 3, pp. 533–553, 2002.

[24] M. Zheng, J. Wang, and M. Q.-H. Meng, "Comparing two gesture design methods for a humanoid robot: Human motion mapping by an RGB-D sensor and hand-puppeteering," presented at the Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on, 2015, pp. 609–614.

[25] M. Hansen and B. Kollmeier, "Continuous assessment of time-varying speech quality," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2888–2899, 1999.

[26] Hamberg, Roelof and de Ridder, Huib, "Continuous assessment of perceptual image quality," *Journal of the Optical Society of America*, vol. 12, no. 12, pp. 2573–2577, 1995.

[27] I. de Kok, "Listening heads," PhD Thesis, University of Twente, Enschede, The Netherlands, 2013.